**SELENIUM ASSIGNMENT**

Q1: Write a python program to scrape data for “Data Analyst” Job position in “Bangalore” location. You have to scrape the job-title, job-location, company\_name, experience\_required. You have to scrape first 10 jobs data. This task will be done in following steps: 1. First get the webpage https://www.shine.com/ 2. Enter “Data Analyst” in “Job title, Skills” field and enter “Bangalore” in “enter the location” field. 3. Then click the searchbutton. 4. Then scrape the data for the first 10 jobs results you get. 5. Finally create a dataframe of the scraped data.

Ans:

from selenium import webdriver

import pandas as pd

driver = webdriver.Chrome('path\_to\_chromedriver')

driver.get('https://www.shine.com/')

Enter the job title and location:

job\_title = driver.find\_element\_by\_id('id\_q')

job\_title.send\_keys('Data Analyst')

location = driver.find\_element\_by\_id('id\_l')

location.send\_keys('Bangalore')

search\_button = driver.find\_element\_by\_xpath('//button[@type="submit"]')

search\_button.click()

job\_titles = driver.find\_elements\_by\_xpath('//a[@class="job\_title\_anchor"]')

job\_locations = driver.find\_elements\_by\_xpath('//li[@class="w-30 mr-10 result-display-location"]/span')

company\_names = driver.find\_elements\_by\_xpath('//a[@class="result-display-company-name"]')

experience\_required = driver.find\_elements\_by\_xpath('//li[@class="w-30 mr-10 result-display-exp"]/span')

data = []

for i in range(10):

job = {

'Job Title': job\_titles[i].text,

'Job Location': job\_locations[i].text,

'Company Name': company\_names[i].text,

'Experience Required': experience\_required[i].text

}

data.append(job)

df = pd.DataFrame(data)

driver.quit()

Q2:Write a python program to scrape data for “Data Scientist” Job position in“Bangalore” location. You have to scrape the job-title, job-location, company\_name. You have to scrape first 10 jobs data. This task will be done in following steps: 1. First get the webpage https://www.shine.com/ 2. Enter “Data Scientist” in “Job title, Skills” field and enter “Bangalore” in “enter thelocation” field. 3. Then click the search button. 4. Then scrape the data for the first 10 jobs results you get. 5. Finally create a dataframe of the scraped data.

Ans:

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Step 1: Get the webpage

url = "https://www.shine.com"

response = requests.get(url)

# Step 2: Enter search criteria and click search button

job\_title = "Data Scientist"

location = "Bangalore"

payload = {

"search\_query": job\_title,

"loc\_query": location

}

response = requests.post(url, data=payload)

# Step 3: Scrape the data for the first 10 jobs

soup = BeautifulSoup(response.content, "html.parser")

job\_results = soup.find\_all("div", class\_="result-display")

job\_data = []

for result in job\_results[:10]:

title = result.find("h2").text.strip()

company = result.find("span", class\_="company-name").text.strip()

location = result.find("span", class\_="location").text.strip()

job\_data.append({"Job Title": title, "Company Name": company, "Location": location})

# Step 4: Create a dataframe of the scraped data

df = pd.DataFrame(job\_data)

# Print the dataframe

print(df)

Q3: In this question you have to scrape data using the filters available on the webpage You have to use the location and salary filter. You have to scrape data for “Data Scientist” designation for first 10 job results. You have to scrape the job-title, job-location, company name, experience required. The location filter to be used is “Delhi/NCR”. The salary filter to be used is “3-6” lakhs The task will be done as shown in the below steps: 1. first get the web page https://www.shine.com/ 2. Enter “Data Scientist” in “Skill, Designations, and Companies” field. 3. Then click the search button. 4. Then apply the location filter and salary filter by checking the respective boxes 5. Then scrape the data for the first 10 jobs results you get. 6. Finally create a dataframe of the scrapeddata.

Ans:

import requests

from bs4 import BeautifulSoup

import pandas as pd

url = "https://www.shine.com/"

response = requests.get(url)

soup = BeautifulSoup(response.content, 'html.parser')

search\_input = soup.find('input', {'id': 'txt\_search'})

search\_input['value'] = 'Data Scientist'

search\_button = soup.find('button', {'id': 'btn\_search'})

response = requests.post(url, data={'txt\_search': 'Data Scientist'})

soup = BeautifulSoup(response.content, 'html.parser')

location\_filter = soup.find('input', {'id': 'chk\_location\_1'})

location\_filter['checked'] = True

salary\_filter = soup.find('input', {'id': 'chk\_salary\_1'})

salary\_filter['checked'] = True

job\_listings = soup.find\_all('div', {'class': 'w-100'})

data = []

for job in job\_listings[:10]:

title = job.find('h3').text.strip()

location = job.find('span', {'class': 'location'}).text.strip()

company = job.find('span', {'class': 'company-name'}).text.strip()

experience = job.find('span', {'class': 'exp'}).text.strip()

data.append([title, location, company, experience])

df = pd.DataFrame(data, columns=['Job Title', 'Job Location', 'Company Name', 'Experience Required'])

Q4: Scrape data of first 100 sunglasses listings on flipkart.com. You have to scrape four attributes: 6. Brand 7. ProductDescription 8. Price The attributes which you have to scrape is ticked marked in the below image

To scrape the data you have to go through following steps: 1. Go to Flipkart webpage by url :https://www.flipkart.com/ 2. Enter “sunglasses” in the search fieldwhere “search for products, brands and more” is written and click the search icon 3. After that you will reach to the page having a lot of sunglasses. From this page you can scrap the required data as usual.

4. After scraping data from the first page, go to the “Next” Button at the bottom other page , then click on it. 5. Now scrape data from this page as usual 6. Repeat this until you get data for 100sunglasses.

Ans:

import requests

from bs4 import BeautifulSoup

url = "https://www.flipkart.com/"

search\_query = "sunglasses"

max\_listings = 100

scraped\_data = []

while len(scraped\_data) < max\_listings:

response = requests.get(url)

soup = BeautifulSoup(response.content, "html.parser")

# Find the search field and enter the search query

search\_field = soup.find("input", attrs={"title": "Search for products, brands and more"})

search\_field["value"] = search\_query

# Click the search icon

search\_icon = soup.find("button", attrs={"type": "submit"})

response = requests.post(url, data=search\_icon.form)

soup = BeautifulSoup(response.content, "html.parser")

# Find the container holding the listings

listings\_container = soup.find("div", attrs={"class": "\_1AtVbE"})

# Extract the required attributes from each listing

for listing in listings\_container.find\_all("div", attrs={"class": "\_2kHMtA"}):

brand = listing.find("div", attrs={"class": "\_2WkVRV"}).text

description = listing.find("a", attrs={"class": "IRpwTa"}).text

price = listing.find("div", attrs={"class": "\_30jeq3 \_1\_WHN1"}).text

scraped\_data.append({"Brand": brand, "ProductDescription": description, "Price": price})

if len(scraped\_data) == max\_listings:

break

# Find the "Next" button and navigate to the next page

next\_button = soup.find("a", attrs={"class": "\_1LKTO3"})

if next\_button:

url = "https://www.flipkart.com" + next\_button["href"]

else:

break

# Print the scraped data

for data in scraped\_data:

print(data)

Q6: Scrape data forfirst 100 sneakers you find whenyou visit flipkart.com and search for “sneakers” inthe search field. You have to scrape 3 attributes of each sneaker: 1. Brand 2. ProductDescription 3. Price As shown in the below image, you have to scrape the above attributes.

Ans:

# let's first connect to the web driver

driver = webdriver.Chrome(r"C:\Users\Arodipa\Downloads\chromedriver\_win32\chromedriver.exe")

url="https://www.flipkart.com/"

driver.get(url)

# finding element for job search bar

search\_g= driver.find\_element\_by\_xpath("//input[@type='text']")

search\_g

# write on search bar

search\_g.send\_keys('sneakers')

search\_btn=driver.find\_element\_by\_xpath("//button[@class='L0Z3Pu']")

search\_btn

search\_btn=driver.find\_element\_by\_class\_name('L0Z3Pu')

search\_btn.click()

B\_name=[]

Price=[]

P\_desc=[]

Discount=[]

for i in range(3):

b\_name=driver.find\_elements\_by\_xpath("//div[@class='\_2WkVRV']")

p\_desc=driver.find\_elements\_by\_xpath("//a[@class='IRpwTa']")

price =driver.find\_elements\_by\_xpath("//div[@class='\_25b18c']")

discount=driver.find\_elements\_by\_xpath("//div[@class='\_3Ay6Sb']")

for j in b\_name:

B\_name.append(j.text)

B\_name[:100]

for k in p\_desc:

P\_desc.append(k.text)

P\_desc[:100]

for l in price:

Price.append(l.text)

Price[:100]

for t in discount:

Discount.append(t.text)

Discount[:100]

print(len(B\_name[:100])),print(len(Price[:100])),print(len(P\_desc[:100])),print(len(Discount[:100]))

Q7: Go to webpage https://www.amazon.in/ Enter “Laptop” in the search field and then click the search icon. Then set CPU Type filter to “Intel Core i7” as shown in the below image:

After setting the filters scrape first 10 laptops data. You have to scrape 3 attributes for each laptop: 1. Title 2. Ratings 3. Price

# let's first connect to the web driver

driver = webdriver.Chrome(r"C:\Users\Arodipa\Downloads\chromedriver\_win32\chromedriver.exe")

url=" https://www.amazon.in "

driver.get(url)

# finding element for job search bar

search\_g= driver.find\_element\_by\_xpath("//input[@type='text']")

search\_g

# write on search bar

search\_g.send\_keys('Laptop')

search\_btn=driver.find\_element\_by\_xpath("//input[@id='nav-search-submit-button']")

search\_btn

search\_btn=driver.find\_element\_by\_xpath("//input[@id='nav-search-submit-button']")

search\_btn.click()

No core filter are shown on the website

Title=[]

Price=[]

Rating=[]

for i in range(3):

b\_name=driver.find\_elements\_by\_xpath("//div[@class='\_2WkVRV']")

p\_desc=driver.find\_elements\_by\_xpath("//a[@class='IRpwTa']")

price =driver.find\_elements\_by\_xpath("//div[@class='\_25b18c']")

for j in b\_name:

Title.append(j.text)

Title[:100]

for k in p\_desc:

P\_desc.append(k.text)

P\_desc[:100]

for l in price:

Price.append(l.text)

Price[:100]

Q8: Write a python program to scrape data for Top 1000 Quotes of All Time. The above task will be done in following steps: 1. First get the webpagehttps://www.azquotes.com/ 2. Click on TopQuotes

from selenium import webdriver

from selenium.webdriver.common.by import By

driver = webdriver.Chrome()

driver.get("https://www.azquotes.com/")

top\_quotes\_button = driver.find\_element(By.LINK\_TEXT, "Top Quotes")

top\_quotes\_button.click()

quotes = driver.find\_elements(By.CSS\_SELECTOR, ".title a")

authors = driver.find\_elements(By.CSS\_SELECTOR, ".author a")

types = driver.find\_elements(By.CSS\_SELECTOR, ".kw-box a")

for quote, author, quote\_type in zip(quotes, authors, types):

print("Quote:", quote.text)

print("Author:", author.text)

print("Type of Quote:", quote\_type.text)

print()

driver.quit()

9: Write a python program to display list of respected former Prime Ministers of India(i.e. Name, Born-Dead, Term of office, Remarks) from https://www.jagranjosh.com/. This task will be done in following steps: 1. First get the webpagehttps://www.jagranjosh.com/ 2. Then You have to click on the GK option 3. Then click on the List of all Prime Ministers of India 4. Then scrap the mentioned data and make theDataFrame.

Ans:

from selenium import webdriver

import pandas as pd

driver = webdriver.Chrome('path\_to\_chromedriver')

driver.get('https://www.jagranjosh.com/')

gk\_option = driver.find\_element\_by\_link\_text('GK')

gk\_option.click()

pm\_option = driver.find\_element\_by\_link\_text('List of all Prime Ministers of India')

pm\_option.click()

data = []

table = driver.find\_element\_by\_xpath('//table[@class="table4"]')

rows = table.find\_elements\_by\_tag\_name('tr')

for row in rows:

cols = row.find\_elements\_by\_tag\_name('td')

if len(cols) == 4:

name = cols[0].text

born\_dead = cols[1].text

term\_of\_office = cols[2].text

remarks = cols[3].text

data.append([name, born\_dead, term\_of\_office, remarks])

df = pd.DataFrame(data, columns=['Name', 'Born-Dead', 'Term of Office', 'Remarks'])

driver.quit()

Q10: Write a python program to display list of 50 Most expensive cars in the world (i.e. Car name and Price) from https://www.motor1.com/ This task will be done in following steps: 1. First get the webpage https://www.motor1.com/ 2. Then You have to type in the search bar ’50 most expensive cars’ 3. Then click on 50 most expensive carsin the world.. 4. Then scrap the mentioned data and make the dataframe.

Ans:

from selenium import webdriver

import pandas as pd

# Step 1: Get the webpage

driver = webdriver.Chrome('path\_to\_chromedriver') # Replace 'path\_to\_chromedriver' with the actual path to your ChromeDriver executable

driver.get('https://www.motor1.com/')

# Step 2: Type in the search bar

search\_bar = driver.find\_element\_by\_id('search-input')

search\_bar.send\_keys('50 most expensive cars')

search\_bar.submit()

# Step 3: Click on the link

link = driver.find\_element\_by\_link\_text('50 Most Expensive Cars in the World')

link.click()

# Step 4: Scrape the data and create a dataframe

car\_names = driver.find\_elements\_by\_xpath('//div[@class="article-content"]/h3')

car\_prices = driver.find\_elements\_by\_xpath('//div[@class="article-content"]/p')

data = []

for name, price in zip(car\_names, car\_prices):

data.append([name.text, price.text])

df = pd.DataFrame(data, columns=['Car Name', 'Price'])

print(df)

driver.quit()